

## Clustering Pelanggan E-Marketplace Shopee Berdasarkan Kategori Produk dan Pendapatan Menggunakan K-Means

Eko Purwanto\*<sup>1</sup>, Asmara Andhini<sup>2</sup>

<sup>12</sup>Universitas Duta Bangsa

<sup>12</sup>Surakarta, Indonesia

Email: <sup>1</sup>[eko\\_purwanto@udb.ac.id](mailto:eko_purwanto@udb.ac.id), <sup>2</sup>[220101007@mhs.udb.ac.id](mailto:220101007@mhs.udb.ac.id)

### Abstract

*Customer clustering is an effective marketing technique to identify the most potential target market. Grouping customers with various characteristics will impact the company's marketing management. This study uses private data obtained through questionnaires distributed to Shopee e-marketplace customers. The data used consists of 296 Shopee customers from the Surakarta area. This dataset shows similarities, so clustering using the K-Means method is necessary. The best clustering result with the parameter  $k=3$  yielded scores of 0.725 and 0.615. The testing reveals three customer clusters on the Shopee e-marketplace with a relationship between product categories and income. The first cluster consists of customers buying beauty products with incomes ranging from 4 to 6 million, the second cluster includes customers with incomes below 2 million, and the third cluster consists of customers with incomes between 2 and 4 million. This study shows that customers consider their income when making product purchases on Shopee.*

**Keywords:** Clustering, Customers, Shopee, Product, Revenue

### Abstraksi

*Klasterisasi pelanggan adalah teknik untuk pemasaran yang lebih efektif dalam menentukan pasar sasaran yang paling potensial. Mengelompokkan pelanggan dengan berbagai karakteristik akan memengaruhi manajemen pemasaran perusahaan. Penelitian ini menggunakan data pribadi yang diperoleh dari distribusi kuesioner kepada responden pelanggan e-marketplace Shopee. Data yang digunakan mencakup 296 pelanggan Shopee di wilayah Surakarta. Dataset ini memiliki kesamaan, sehingga perlu dilakukan klasterisasi menggunakan metode K-Means. Hasil klasterisasi terbaik dengan parameter  $k=3$  menghasilkan skor 0,725 dan 0,615. Pengujian menunjukkan adanya tiga klaster pelanggan di e-marketplace Shopee dengan hubungan antara kategori produk dan pendapatan. Klaster pertama adalah pelanggan yang membeli produk kecantikan dengan pendapatan antara 4 hingga 6 juta, klaster kedua adalah pelanggan dengan pendapatan kurang dari 2 juta, dan klaster ketiga adalah pelanggan dengan pendapatan antara 2*

hingga 4 juta. Penelitian ini menunjukkan bahwa pelanggan menghitung pendapatan yang dimiliki saat membeli produk di Shopee.

**Kata Kunci:** *Klasterisasi, Pelanggan, Shopee, Produk, Pendapatan*

## **1. PENDAHULUAN**

Klasterisasi pelanggan adalah metode yang efektif untuk mengidentifikasi pasar sasaran yang paling potensial. Mengelompokkan pelanggan berdasarkan karakteristik yang serupa dapat memberikan wawasan mendalam yang berguna bagi manajemen pemasaran perusahaan. Pelanggan merupakan faktor utama dalam kemajuan ekonomi, karena mereka adalah pihak yang langsung terlibat dalam transaksi ekonomi yang mendukung pertumbuhan perusahaan dan perekonomian secara keseluruhan.

E-marketplace, yang merupakan platform untuk jual beli barang secara daring, menjadi sarana utama dalam perdagangan modern. Dalam model Business to Customer (B2C), penjual langsung menjual barang kepada pelanggan melalui e-marketplace [1]. Dengan semakin meningkatnya jumlah pedagang yang bersaing di e-marketplace, penting bagi penjual untuk memiliki strategi pemasaran yang tepat agar dapat mempertahankan dan menarik pelanggan. Salah satu cara yang dapat digunakan untuk mempertahankan potensi pelanggan adalah dengan menerapkan teknologi promosi yang efektif [2]–[5].

Klasterisasi pelanggan menawarkan solusi untuk mengetahui preferensi pasar dengan lebih baik. Dengan mengelompokkan pelanggan berdasarkan karakteristik serupa, penjual dapat memperoleh peluang bisnis yang lebih relevan dan akurat [6]. Selain itu, klasterisasi pelanggan dapat membantu memahami perilaku konsumen, sehingga memberikan akurasi yang lebih baik dalam menyusun rekomendasi produk. Hal ini sangat penting dalam menjaga loyalitas pelanggan, yang merupakan hal utama bagi penjual untuk memastikan pelanggan tetap setia. Salah satu cara untuk menjaga loyalitas pelanggan adalah dengan memberikan perlakuan yang berbeda kepada setiap segmen pelanggan sesuai dengan karakteristik mereka [7].

Metode K-Means merupakan salah satu algoritma yang umum digunakan dalam klasterisasi. Algoritma ini mengelompokkan data menjadi beberapa klaster berdasarkan kesamaan, di mana setiap klaster memiliki jarak kuadrat minimum dari titik pusat klaster (centroid) [8]–[13]. Dalam konteks ini, klasterisasi pelanggan bertujuan untuk mengelompokkan data pelanggan berdasarkan karakteristik yang serupa [14]–[16], yang pada gilirannya dapat meningkatkan relevansi rekomendasi produk serta meningkatkan retensi pelanggan dan keuntungan penjual [8]–[13], [17].

Shopee, salah satu e-marketplace terbesar di Asia Tenggara pada tahun 2020, telah menjadi platform yang sangat populer. Shopee memberikan berbagai promo menarik seperti diskon dan pengiriman gratis, yang menjadi daya tarik utama bagi banyak pelanggan yang melakukan transaksi daring [18].

Selain itu, pemahaman tentang demografi pelanggan juga sangat penting. Demografi adalah studi mengenai karakteristik populasi manusia, yang mencakup usia,

jenis kelamin, status pernikahan, dan pekerjaan. Analisis demografis membantu dalam memahami segmen pasar dan bagaimana karakteristik tersebut memengaruhi preferensi dan perilaku pelanggan [19]. Oleh karena itu, pengelompokan pelanggan di e-marketplace berdasarkan karakteristik demografis ini sangat penting untuk merancang strategi pemasaran yang lebih efektif.

Penelitian ini bertujuan untuk melakukan analisis pengelompokan pelanggan Shopee berdasarkan karakteristik demografis mereka. Dalam penelitian ini, analisis klasterisasi pelanggan akan menggunakan algoritma K-Means untuk mengidentifikasi segmen-segmen pasar yang relevan dan memberikan wawasan bagi pengembangan strategi pemasaran yang lebih terfokus.

## **2. TINJAUAN PUSTAKA**

Klasterisasi pelanggan merupakan teknik yang banyak digunakan dalam pemasaran untuk mengelompokkan pelanggan berdasarkan karakteristik yang serupa. Teknik ini bertujuan untuk membantu perusahaan merancang strategi pemasaran yang lebih efektif dengan memahami perilaku pelanggan [20], [21]. Beberapa penelitian sebelumnya telah mengungkapkan bahwa klasterisasi pelanggan dapat memberikan wawasan yang sangat berguna dalam merancang kampanye pemasaran yang lebih terarah. Salah satu metode klasterisasi yang paling populer adalah K-Means, sebuah algoritma yang mengelompokkan data berdasarkan kedekatannya terhadap pusat klaster (centroid) [22].

Meskipun K-Means efektif untuk mengelola data dalam jumlah besar, algoritma ini memiliki kekurangan, terutama ketergantungannya pada pemilihan jumlah klaster ( $k$ ) yang tepat. Selain itu, K-Means juga sensitif terhadap keberadaan outlier [23], yang dapat memengaruhi hasil klasterisasi, terutama ketika ada variasi ekstrem dalam data pelanggan. Beberapa penelitian menunjukkan bahwa meskipun K-Means mampu memberikan hasil yang baik dalam klasterisasi, namun pemilihan  $k$  yang tidak tepat dapat mengurangi akurasi klaster yang terbentuk [24], [25].

Selain itu, analisis demografis juga memainkan peranan penting dalam klasterisasi pelanggan. Pemahaman tentang karakteristik demografis pelanggan, seperti usia, jenis kelamin, dan status sosial, sangat berguna bagi perusahaan e-marketplace untuk menyusun strategi pemasaran yang lebih terarah dan relevan bagi masing-masing segmen pelanggan. Berbagai studi menunjukkan bahwa integrasi data demografis dalam klasterisasi dapat memperkaya hasil analisis dan memberikan pemahaman yang lebih mendalam tentang preferensi pelanggan.

Berdasarkan penelitian terdahulu, K-Means telah terbukti efektif dalam mengelompokkan pelanggan berdasarkan kategori produk yang dibeli dan tingkat pendapatan mereka. Segmentasi ini sangat bermanfaat dalam merancang strategi promosi yang lebih tepat sasaran [26], [27], karena pelanggan dengan karakteristik yang serupa cenderung merespon promosi dengan cara yang sama. Selain itu, loyalitas pelanggan menjadi faktor penting dalam dunia e-commerce. Segmentasi yang akurat

dapat meningkatkan retensi pelanggan dengan menawarkan produk yang lebih relevan dan sesuai dengan kebutuhan mereka [28], [29].

Penelitian yang menggunakan K-Means pada platform e-marketplace, seperti Shopee, memungkinkan untuk melakukan analisis yang lebih mendalam terhadap kelompok pelanggan berdasarkan preferensi produk dan pendapatan. Hal ini memberi penjual wawasan lebih jelas tentang segmen pasar yang ada, sehingga mereka dapat menyusun strategi pemasaran yang lebih efektif [27], [30]. Dalam konteks Shopee, yang menawarkan berbagai promo dan diskon, segmentasi yang tepat dapat membantu meningkatkan konversi penjualan dengan menarik perhatian pelanggan yang tepat.

### 3. METODE PENELITIAN

Metode data mining yang digunakan dalam penelitian ini terdiri dari beberapa tahap: tahap pertama adalah pengumpulan data, tahap kedua adalah pra-pemrosesan, tahap ketiga adalah pemodelan, dan tahap keempat adalah analisis.



Gambar 1. Tahapan Penelitian

#### 3.1. Pengumpulan Data

Proses pengumpulan data dilakukan dengan menyebarkan kuesioner kepada pelanggan Shopee di wilayah Surakarta. Target responden dalam penelitian ini adalah 296 pelanggan e-marketplace Shopee yang telah melakukan transaksi di platform tersebut. Metode pengambilan sampel yang digunakan adalah purposive sampling, yaitu pemilihan responden berdasarkan kriteria tertentu, yaitu pelanggan yang telah melakukan pembelian produk di Shopee dalam tiga bulan terakhir. Dengan jumlah responden tersebut, penelitian ini telah memenuhi syarat minimum untuk melakukan klusterisasi menggunakan metode K-Means. Jumlah sampel yang digunakan sudah memadai untuk memastikan hasil klusterisasi yang representatif. Kuesioner yang digunakan dalam penelitian ini terdiri dari beberapa bagian yang mencakup variabel demografis dan perilaku pembelian pelanggan. Variabel yang ditanyakan dalam kuesioner meliputi:

- Demografi: usia, jenis kelamin, pekerjaan, tingkat pendapatan, status pernikahan.
- Perilaku Pembelian: kategori produk yang dibeli, frekuensi pembelian, jumlah pengeluaran per pembelian, dan alasan memilih produk tertentu.

Data yang dikumpulkan akan digunakan untuk mengelompokkan pelanggan berdasarkan preferensi produk dan pendapatan, serta untuk memahami hubungan antara faktor demografis dan perilaku pembelian.

### 3.2. Pra-pemrosesan

Tahap pra-pemrosesan terdiri dari:

1. Pembersihan Data

Tahap ini untuk memeriksa data yang kosong atau tidak mengandung nilai. Data kosong diisi dengan nilai yang sesuai dengan tipe data tersebut. Tahap ini menghapus data yang tidak relevan.

2. Mencari Korelasi antara Fitur

Pada tahap ini, dihitung korelasi antar kolom fitur secara bersamaan.

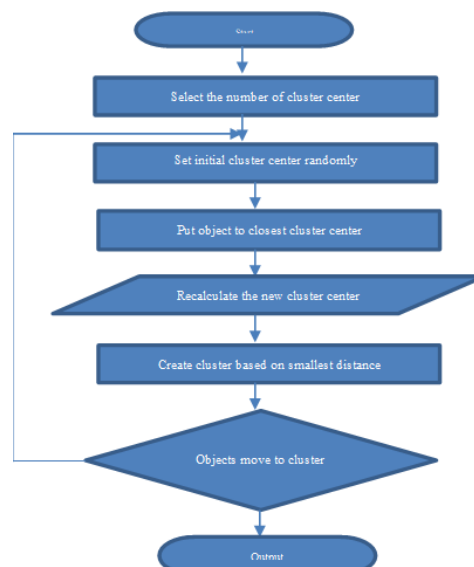
3. Menangani Outliers

Pada tahap ini, beberapa teknik digunakan untuk mempersiapkan data sebelum dilakukan analisis lebih lanjut. Salah satu teknik yang diterapkan adalah menangani outliers dengan menggunakan z-score. Outliers diidentifikasi dengan cara menghitung nilai z-score untuk setiap nilai dalam data, di mana nilai z-score lebih besar dari 3 ( $z > 3$ ) dianggap sebagai outlier dan akan dipertimbangkan untuk dihapus dari dataset. Threshold yang digunakan untuk mendeteksi outliers adalah nilai  $z > 3$ , yang merupakan batas umum yang digunakan dalam banyak penelitian untuk memastikan data yang digunakan memiliki kualitas yang baik.

4. Standarisasi Data

Standarisasi fitur menghapus rata-rata dan mengubahnya menjadi variansi unit. Standarisasi data adalah persyaratan umum untuk banyak estimator machine learning.

### 3.3. Tahapan Klasterisasi



Gambar 2. Diagram Alir Tahapan Pengelompokan

Gambar 2 merupakan tahapan pengelompokkan apada penelitian ini membentuk model menggunakan algoritma K-Means. K-Means adalah algoritma sederhana yang mengelompokkan sejumlah besar objek ke dalam atribut tertentu ke dalam K cluster [31].

Algoritma K-Means dapat membentuk model clustering untuk memetakan pelanggan. Hal ini dilakukan untuk menemukan kelompok pelanggan yang memiliki potensi keuntungan bagi perusahaan [31]. K-Means adalah metode pengelompokan data non-hierarkis dengan menggunakan metode partisi berbasis centroid. Ini adalah cara untuk mengelompokkan data menjadi satu atau lebih [32]. Centroid adalah nilai yang digunakan untuk menghitung jarak dalam data dengan menentukan nilai awal dan nilai untuk setiap iterasi secara acak menggunakan rumus [33].

Pada tahap klasterisasi menggunakan algoritma K-Means, pemilihan jumlah klaster (k) merupakan langkah krusial dalam menentukan hasil klasterisasi yang optimal. Dalam penelitian ini, dilakukan pengujian dengan tiga nilai k yang berbeda: k=3, k=5, dan k=7. Pemilihan k=3, k=5, dan k=7 didasarkan pada pendekatan pengujian yang sering digunakan dalam literatur untuk menentukan jumlah klaster yang optimal, dengan menguji variasi hasil berdasarkan jumlah klaster yang berbeda. Pemilihan nilai k ini akan dibandingkan berdasarkan dua indikator evaluasi, yaitu Silhouette Score dan Davies-Bouldin Index, yang digunakan untuk menilai kualitas klaster yang terbentuk.

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj} \quad (1)$$

Menentukan jarak dari titik Centroid ke titik setiap objek.

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (2)$$

1. Menentukan anggota klaster dengan mengelompokkan objek berdasarkan perhitungan jarak minimum.
2. Pengulangan ke tahap 2 berlanjut hingga kita mendapatkan nilai sentroid dan anggota klaster yang tetap.

### 3.4. Analisis Data

Analisis data digunakan untuk melihat data visualisasi dari pemrosesan dengan algoritma K-Means. Hasil analisis data menjadi panduan untuk menentukan strategi pemasaran [34].

## 4. HASIL DAN PEMBAHASAN

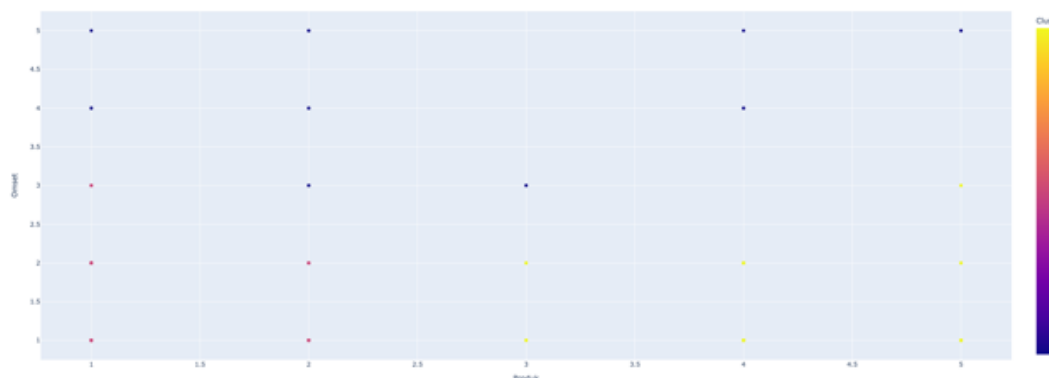
### 4.1. Hasil

Penelitian ini melakukan tahap pengujian kinerja dari metode yang diusulkan. Pengujian K-Means dilakukan dengan mengubah nilai parameter k (k=3, k=5, dan k=7). Pengujian parameter juga dilakukan pada dua fitur atau kolom yang berbeda: Produk dan Pendapatan. Hasil pengujian dengan parameter k=3 menunjukkan adanya 3 klaster dengan data pada Tabel 1 dan Gambar 3.

Tabel 1. Klaster Produk dan Pendapatan dengan parameter k=3

Data	Gender	Old	Products	Revenue	Cluster
0	2	2	1	1	1
1	2	2	1	1	1
2	1	4	1	3	1
3	2	2	1	1	1

Data	Gender	Old	Products	Revenue	Cluster
4	1	4	2	5	0
...	...	...	...	...	...
281	2	2	1	1	1
282	2	2	2	1	1
283	1	2	1	1	1
284	2	3	1	1	1

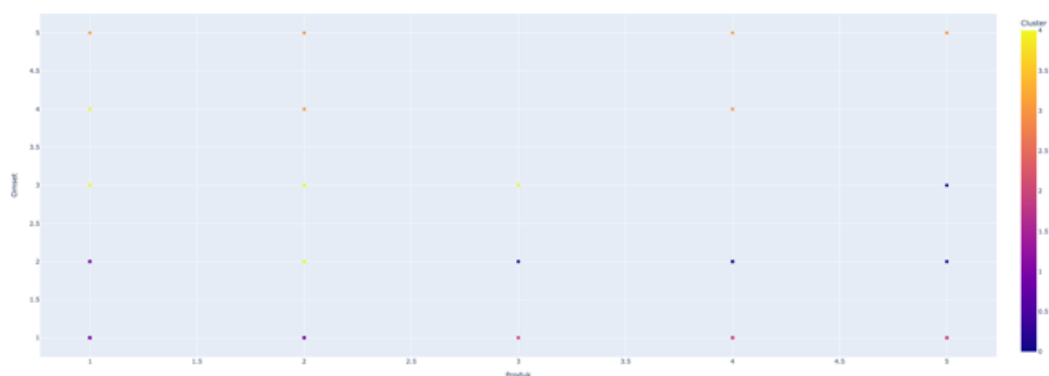


Gambar 3. Hasil Cluster Hubungan Produk dan Pendapatan dengan Parameter K-3

Kinerja K-Means dengan parameter  $k=3$  menghasilkan skor Silhouette sebesar 0,725 dan skor Davies Bouldin sebesar 0,615. Hasil pengujian dengan parameter  $k=5$  menunjukkan 5 cluster dengan data pada Tabel 2 dan Gambar 4.

Tabel 2. Klaster Produk dan Pendapatan dengan parameter  $k=5$

Data	Gender	Old	Products	Revenue	Cluster
0	2	2	1	1	1
1	2	2	1	1	1
2	1	4	1	3	4
3	2	2	1	1	1
4	1	4	2	5	3
...	...	...	...	...	...
281	2	2	1	1	1
282	2	2	2	1	1
283	1	2	1	1	1
284	2	3	1	1	1

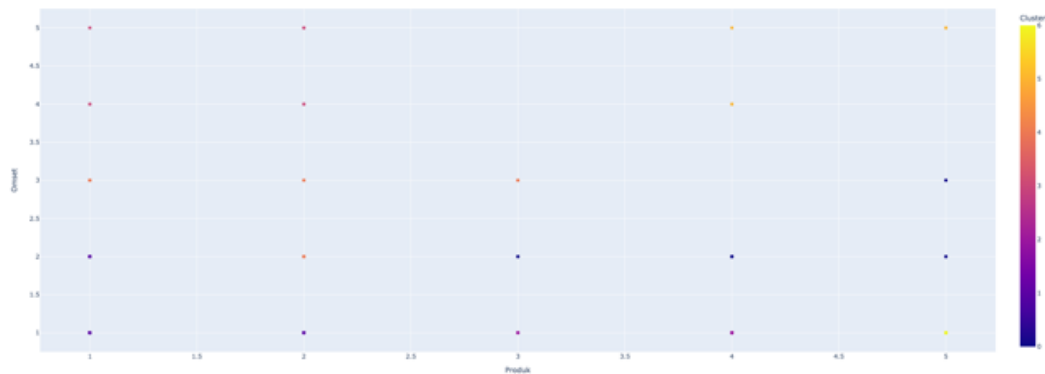


Gambar 4. Hasil Klaster Hubungan Produk dan Pendapatan dengan Parameter K-5

Kinerja K-Means dengan parameter k=5 menghasilkan skor Silhouette sebesar 0,602 dan skor Davies Bouldin sebesar 0,943. Hasil pengujian dengan parameter k=7 menunjukkan 7 kluster dengan data pada Tabel 3 dan Gambar 5.

Tabel 3. Kluster Produk dan Pendapatan dengan parameter k-5

Data	Gender	Old	Products	Revenue	Cluster
0	2	2	1	1	1
1	2	2	1	1	1
2	1	4	1	3	4
3	2	2	1	1	1
4	1	4	2	5	3
...	...	...	...	...	...
281	2	2	1	1	1
282	2	2	2	1	1
283	1	2	1	1	1
284	2	3	1	1	1



Gambar 5. Hasil Kluster Hubungan Produk dan Pendapatan dengan Parameter K-7  
Kinerja K-Means dengan parameter k=7 menghasilkan skor Silhouette sebesar 0,699 dan skor Davies Bouldin sebesar 0,670.

#### 4.2. Pembahasan

Kinerja metode K-Means kemudian divisualisasikan dalam bentuk grafik. Pengukuran kinerja dapat dilihat dari Skor Silhouette dan Skor Davies-Bouldin.

Tabel 4. Hasil Kluster Hubungan Produk dan Pendapatan

	k=3	k=5	k=7
Silhouette score	0.725	0.602	0.699
Davies bouldin score	0.615	0.943	0.670

Pengujian kluster dilakukan terhadap hubungan antara produk dan pendapatan. Performa K-Means dengan parameter k=3 menghasilkan skor terbaik, yaitu 0,725 dan 0,615.



Hasil pengujian menunjukkan tiga klaster pelanggan di e-marketplace Shopee dengan hubungan antara produk dan omzet. Pelanggan menghitung pendapatan yang mereka miliki saat membeli produk di e-marketplace Shopee.

Produk dibagi menjadi lima kategori: fesyen, kecantikan, makanan, minuman dan bahan makanan, serta peralatan rumah tangga. Omzet penjualan dibagi menjadi lima kategori, yaitu >Rp. 2.000.000, Rp. 2.000.001-Rp. 4.000.000, Rp. 4.000.001-Rp. 6.000.000, Rp. 6.000.001-Rp. 8.000.000, dan <Rp. 8.000.000.

Hasil penelitian ini menunjukkan adanya tiga klaster pelanggan di e-marketplace Shopee. Klaster pertama adalah pelanggan yang membeli produk kecantikan dengan pendapatan 4 hingga 6 juta rupiah. Klaster kedua adalah pelanggan yang membeli produk kecantikan dengan pendapatan kurang dari 2 juta rupiah. Sementara itu, klaster ketiga adalah pelanggan yang membeli produk kecantikan dengan pendapatan 2 hingga 4 juta rupiah.

Penelitian ini menunjukkan bahwa produk kecantikan yang dijual di e-marketplace Shopee sangat diminati oleh pelanggan dengan pendapatan antara 2 hingga 6 juta rupiah.

## **5. KESIMPULAN**

Penelitian ini bertujuan untuk mengidentifikasi segmen pelanggan e-marketplace Shopee di Surakarta melalui klasterisasi menggunakan metode K-Means, berdasarkan kategori produk yang dibeli dan pendapatan pelanggan. Hasil analisis menunjukkan bahwa terdapat tiga klaster utama pelanggan Shopee di Surakarta, masing-masing dengan karakteristik yang berbeda. Klaster pertama terdiri dari pelanggan dengan pendapatan antara 4 hingga 6 juta rupiah yang cenderung membeli produk kecantikan. Klaster kedua melibatkan pelanggan dengan pendapatan lebih rendah, yang juga memiliki preferensi yang sama terhadap produk kecantikan. Sementara itu, klaster ketiga melibatkan pelanggan dengan pendapatan antara 2 hingga 4 juta rupiah yang membeli produk kecantikan.

Namun, penelitian ini memiliki beberapa keterbatasan. Pertama, penelitian hanya berfokus pada wilayah Surakarta, yang membatasi generalisasi temuan kepada wilayah lain di Indonesia atau pasar internasional. Kedua, sampel yang digunakan terbatas pada 296 responden, yang mungkin tidak sepenuhnya mewakili keseluruhan populasi pelanggan Shopee. Selain itu, penelitian ini hanya mempertimbangkan dua variabel, yaitu kategori produk dan pendapatan. Variabel lain yang mungkin mempengaruhi perilaku pelanggan, seperti usia, jenis kelamin, atau frekuensi pembelian, tidak dianalisis dalam penelitian ini.

Saran untuk penelitian selanjutnya adalah untuk memperluas cakupan penelitian dengan menggunakan sampel yang lebih besar dan mencakup wilayah yang lebih luas atau bahkan secara internasional untuk memperoleh hasil yang lebih representatif. Penelitian berikutnya juga bisa mempertimbangkan faktor-faktor lain yang mempengaruhi perilaku pembelian pelanggan, seperti preferensi pribadi, pengaruh

media sosial, atau kebiasaan belanja online lainnya. Selain itu, penggunaan metode klusterisasi lain seperti DBSCAN atau Hierarchical Clustering dapat menjadi alternatif untuk membandingkan hasil klusterisasi yang diperoleh, yang bisa memberikan wawasan tambahan tentang segmentasi pelanggan.

Dari temuan yang ada, implikasi praktis yang dapat diambil adalah bahwa Shopee dan penjual di platform ini dapat menggunakan hasil klusterisasi untuk menyesuaikan strategi pemasaran mereka. Misalnya, bagi klaster dengan pendapatan lebih tinggi, Shopee dapat menawarkan produk premium atau eksklusif, serta promosi dengan harga yang lebih tinggi, sementara klaster dengan pendapatan lebih rendah dapat diberikan penawaran yang lebih terjangkau atau promosi diskon khusus. Dengan memahami karakteristik masing-masing klaster pelanggan, Shopee dapat meningkatkan retensi pelanggan dan konversi penjualan, serta menyediakan penawaran yang lebih personal dan relevan.

Secara keseluruhan, penelitian ini memberikan wawasan penting bagi strategi pemasaran berbasis data di e-marketplace Shopee dan dapat menjadi dasar bagi pengembangan lebih lanjut dalam segmentasi pasar yang lebih spesifik dan tepat sasaran.

#### DAFTAR PUSTAKA

- [1] F. Alzami *et al.*, "Implementation Of ETL E-Commerce For Customer Clustering Using RFM And K-Means Clustering," *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, vol. 10, no. 3, p. 167, 2022, doi: 10.24843/jim.2022.v10.i03.p05.
- [2] M. Mardalius and T. Christy, "Mapping of Potential Customers As a Clothing Promotion Strategy Using K-Means Clustering Algorithm," ... *J. Ilmu Pengetah. Dan ...*, vol. 6, no. 1, pp. 67–72, 2020, doi: 10.33480/jitk.v6i1.1414.
- [3] R. Gilton, V. V. Mugobo, and W. Jooste, "The Effectiveness of E-Marketplaces as Trading Platforms in Cape Town," *Expert J. Mark.*, vol. 11, no. 1, pp. 78–90, 2023, [Online]. Available: <http://marketing.expertjournals.com78>
- [4] E. Maier and J. Wieringa, "Acquiring customers through online marketplaces? The effect of marketplace sales on sales in a retailer's own channels," *Int. J. Res. Mark.*, vol. 38, no. 2, pp. 311–328, 2021, doi: 10.1016/j.ijresmar.2020.09.007.
- [5] D. K. Shetty, R. K. Raju, N. Ayedee, and B. Singla, "Assessment of E-Marketplace in Increasing the Cost Efficiency of Road Transport Industry," *PalArch's J. ...*, vol. 17, no. 9, pp. 3799–3840, 2020, [Online]. Available: <https://archives.palarch.nl/index.php/jae/article/view/4508>
- [6] F. Zheng and Q. Liu, "Anomalous Telecom Customer Behavior Detection and Clustering Analysis Based on ISP's Operating Data," *IEEE Access*, vol. 8, pp. 42737–42738, 2020, doi: 10.1109/ACCESS.2020.2976898.
- [7] U. Firdaus and D. N. Utama, "Development of bank's customer segmentation model based on rfm+b approach," *ICIC Express Lett. Part B Appl.*, vol. 12, no. 1, pp. 17–26, 2021, doi: 10.24507/icicelb.12.01.17.
- [8] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [9] T. R. Kurniawan, E. Chumaidiyah, and L. Andrawina, "Application of the k-means clustering method and simple linear regression to new student admissions as a promotion method," *J. Infotel*, vol. 15, no. 1, pp. 25–33, 2023, doi:

- 10.20895/infotel.v15i1.860.
- [10] I. B. G. Sarasvananda, R. Wardoyo, and A. K. Sari, "The K-Means Clustering Algorithm With Semantic Similarity To Estimate The Cost of Hospitalization," *IJCCS (Indonesian J. Comput. Cybern. Syst.,* vol. 13, no. 4, p. 313, 2019, doi: 10.22146/ijccs.45093.
  - [11] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J*, vol. 2, no. 2, pp. 226–235, 2019, doi: 10.3390/j2020016.
  - [12] P. Govender and V. Sivakumar, *Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)*, vol. 11, no. 1. Turkish National Committee for Air Pollution Research and Control, 2020. doi: 10.1016/j.apr.2019.09.009.
  - [13] L. Fitriyani, N. Komariah, Y. Bintang Fajar, S. Saepudin, and M. Muslih, "Application of K-Means Algorithm on Favorite Menu Selection At Warung Nasi Alam Sunda," *Int. J. Eng. Appl. Technol.*, vol. 5, no. 1, pp. 1–16, 2022, doi: 10.52005/ijeat.v5i1.55.
  - [14] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 336, no. 1, 2018, doi: 10.1088/1757-899X/336/1/012017.
  - [15] A. Abdulhafedh, "Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation," *J. City Dev.*, vol. 3, no. 1, pp. 12–30, 2021, doi: 10.12691/jcd-3-1-3.
  - [16] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, "RFM ranking – An effective approach to customer segmentation," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 33, no. 10, pp. 1251–1257, 2021, doi: 10.1016/j.jksuci.2018.09.004.
  - [17] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *Sustain.*, vol. 14, no. 12, pp. 1–15, 2022, doi: 10.3390/su14127243.
  - [18] N. M. Ulfa and A. F. Indriya Himawan, "The Effect of Celebrity Endorsers, Service Features, and Promotions for Shopee E-marketplace Purchase Interest," *Indones. Vocat. Res. J.*, vol. 2, no. 1, p. 30, 2022, doi: 10.30587/ivrj.v2i1.4780.
  - [19] V. Kónya, D. Matić, and J. Pavlović, "The influence of demographics, job characteristics and characteristics of organizations on employee commitment," *Acta Polytech. Hungarica*, vol. 13, no. 3, pp. 119–138, 2016, doi: 10.12700/aph.13.3.2016.3.7.
  - [20] L. Al-Dabbas, H. Al-Tarawneh, and T. A. Al-Rawashdeh, "Customer Personality Segmentation Using K-Means Clustering," in *2023 International Conference on Information Technology (ICIT)*, 2023, pp. 537–543. doi: 10.1109/icit58056.2023.10225996.
  - [21] P. Perapu, "Customer Segmentation Using K-Means Clustering for Personalized Marketing Campaigns," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 11, no. 3, pp. 810–815, 2025, doi: 10.32628/cseit25113344.
  - [22] D. Yu, S. Dong, and S. Yao, "Improvement of K-Means Algorithm and Its Application in Air Passenger Grouping," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–13, 2022, doi: 10.1155/2022/3958423.
  - [23] J. Zhao, Y. Bao, D. Li, and X. Guan, "An Improved K-Means Algorithm Based on Contour Similarity," *Mathematics*, vol. 12, no. 14, p. 2211, 2024, doi: 10.3390/math12142211.

- [24] N. N. L. Rifqiah, M. G. Hamid, and E. R. Kamila, "Analisis Perilaku Konsumen dan Pembelian Impulsif Pada E-commerce Shopee," *J. Nuansa*, vol. 3, no. 1, pp. 35–43, 2025, doi: 10.61132/nuansa.v3i1.1506.
- [25] D. Abhichandani, N. R. T. Vadrevu, P. Doshi, and S. Shrivastava, "Predicting Online Purchases Using Six Machine Learning Models Based on Customer Demographics," in *2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*, 2025, pp. 1787–1792. doi: 10.1109/icdici66477.2025.11135228.
- [26] S. N. Lathifah and Z. F. Azzahra, "AI-Driven Customers Segmentation Using K-Means Clustering," *G-Tech*, vol. 9, no. 1, pp. 320–329, 2025, doi: 10.70609/gtech.v9i1.6202.
- [27] M. Hamza, H. Haryadi, and S. Sularno, "Segmentasi Pelanggan Berdasarkan Pendapatan Dan Skor Pengeluaran Menggunakan Metode K-Means," *J. Sains dan Teknol.*, vol. 4, no. 3, pp. 325–337, 2024, doi: 10.47233/jsit.v4i3.2468.
- [28] F. Irawan, F. Amaluddin, and A. Wijayanti, "E-Commerce Customer Segmentation Based on RFM Analysis Using DBSCAN Algorithm to Improve Marketing Strategy," *J. EMT KITA*, vol. 9, no. 4, pp. 1850–1859, 2025, doi: 10.35870/emt.v9i4.5375.
- [29] R. Fitriyani, A. L. Firmansyah, A. Y. N. Fithri, and L. A. Nurfadillah, "Penerapan Algoritma Clustering untuk Segmentasi Pelanggan E-commerce berdasarkan Data Pembelian dan Aktivitas," *Semin. Teknol. MAJALENGKA*, vol. 8, pp. 372–379, 2024, doi: 10.31949/stima.v8i0.1129.
- [30] R. M. Fauzan and G. Alfian, "Segmentasi Pelanggan E-Commerce Menggunakan Fitur Recency, Frequency, Monetary (RFM) dan Algoritma Klasterisasi K-Means," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 9, no. 3, pp. 170–177, 2024, doi: 10.14421/jiska.2024.9.3.170-177.
- [31] D. A. T. Devanta, "Optimization of the K-Means Clustering Algorithm Using Davies Bouldin Index in Iris Data Classification," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 545–552, 2023, doi: 10.30865/klik.v4i1.964.
- [32] K. B, "A Comparative Study on K-Means Clustering and Agglomerative Hierarchical Clustering," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 5, pp. 1600–1604, 2020, doi: 10.30534/ijeter/2020/20852020.
- [33] E. Umargono, J. E. Suseno, and V. G. S. K., "K-Means Clustering Optimization using the Elbow Method and Early Centroid Determination Based-on Mean and Median," vol. 474, no. Isstec 2019, pp. 234–240, 2020, doi: 10.5220/0009908402340240.
- [34] S. Abadi *et al.*, "Application model of k-means clustering: Insights into promotion strategy of vocational high school," *Int. J. Eng. Technol.*, vol. 7, no. 2.27 Special Issue 27, pp. 182–187, 2018, doi: 10.14419/ijet.v7i2.11491.